

ガウス型基底関数を用いた密度汎関数法における 分子積分関与ルーチンの高速化

田原才静 吉廣 保 佐藤文俊

東京大学生産技術研究所 計算科学技術連携研究センター

1. はじめに

ProteinDF はガウス型基底関数を用いた密度汎関数法に基づく大規模タンパク質の量子化学計算ソフトウェアであり[1]、104 残基の金属タンパク質シクロム *c* の全電子計算に成功している[2]。

しかしその処理速度にはまだ改良の余地がある。特に Kohn-Sham 行列と全エネルギーの計算に必要な分子積分関与ルーチンは、100 残基程度のタンパク質を計算する場合に全計算時間の約半分を占めており計算律速となっている。

タンパク質の全電子計算は、タンパク質の基礎過程や機能の解明に重要な役割を示すため、これからますます需要が高まることが予想される。そこで本研究は、100 残基程度のタンパク質全電子計算の律速である分子積分関与ルーチンを、現在の計算機メモリ事情にあわせて高速化することを目的とする。

2. 現在の ProteinDF

ProteinDF はオブジェクト指向言語 C++ でコーディングされている。現バージョンの ProteinDF では、Kohn-Sham 行列を求める DfFockmatrix クラスと全エネルギーを求める DfTotalenergy クラスにおいて、1 回ずつ 3 中心の分子積分計算を行う DfEri クラスと DfOverlap クラスを共に呼び出している。また分子軌道から全電子密度を計算する DfDensityfitting クラスが 1 回 DfEri を呼び出しており、合わせて 5 回分子積分計算を行うクラスを呼び出している(図1)。

ところが以下の式に示すように、分子積分部は Kohn-Sham 行列計算と全エネルギー計算とで共通部分が多い。それにも関わらず別々に分子積分計算クラスを呼び出しているのは、Kohn-Sham 行列と全エネルギーの計算では分子積分と内積をとる展開係数が異なる (μ_γ 、 ε_γ) ことに

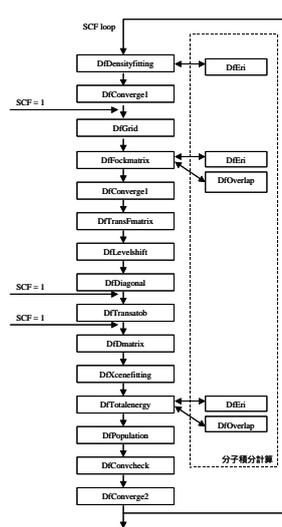


図1 現行ProteinDFにおけるSCF処理の流れ

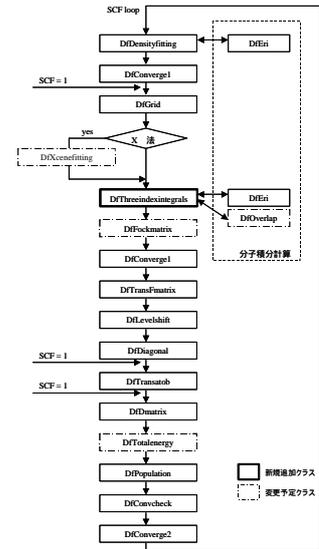


図2 高速化後ProteinDFにおけるSCF処理の流れ

より、これらを同時に計算するには多くのメモリが必要であるためである。

(1) Kohn-Sham 行列作成式

$$F_{pq} = h_{pq} + \sum_{\alpha} \rho_{\alpha} \langle pq | \alpha \rangle + \sum_{\gamma} \mu_{\gamma} \langle pqr \rangle$$

(2) 全エネルギー計算式

$$E = \sum_{pq} P_{pq} \left\{ h_{pq} + \sum_{\alpha} \rho_{\alpha} \langle pq | \alpha \rangle + \sum_{\gamma} \varepsilon_{\gamma} \langle pqr \rangle \right\} - \frac{1}{2} \sum_{\alpha\beta} \rho_{\alpha} \rho_{\beta} \langle \alpha | \beta \rangle + (NN)$$

High-speed calculation of the molecular-integral routines in density-functional method with gaussian type basis sets.

Saisei Tahara, Tamotsu Yoshihiro, Fumitoshi Sato
Institute of industrial Science, University of Tokyo

3. 高速化の手法

コンピュータの著しい発展により、以前に比べて多量のメモリを利用できるようになっている。そこでこれまで不可能であった、分子積分が関与する計算をあらかじめまとめて行い計算回数を減少させる手法により、ProteinDF の高速化を行った。

また、ProteinDF では SCF 繰り返し計算において、1 回前の計算結果を利用するアップデート法を採用している。分子積分を別々に計算する現在の方法では、全エネルギーの計算にこれを用いても高速化できないが、本方法ではアップデート法を併用して高速化することが可能となり、更なる高速化を図ることができた。

3 - 1 多中心分子積分クラスの追加

図 2 に示すように新しいクラス DfThreeindexintegrals を DfFockmatrix の前に作成した。DfThreeindexintegrals は、DfEri と DfOverlap を呼び出し、 $\sum_{\alpha} \rho_{\alpha} \langle pq | \alpha \rangle$ 、 $\sum_{\gamma} \mu_{\gamma} \langle pqr \rangle$ および $\sum_{\gamma} \varepsilon_{\gamma} \langle pqr \rangle$ の計算結果をファイルに出力する。そしてこの計算結果を利用して Kohn-Sham 行列(1)と全エネルギー(2)とを求め。

この変更によって、分子積分クラスの呼び出し回数が現行の 3/5 回となりアップデート法を併用すると 2 倍程度の高速化が見込まれる。

本クラスでは SCF=1 のときに、Kohn-Sham 行列作成のため、

$$\sum_{\alpha} \rho_{\alpha} \langle pq | \alpha \rangle + \sum_{\gamma} \mu_{\gamma} \langle pqr \rangle \text{ を}$$

全エネルギー計算のため、

$$\sum_{\alpha} \rho_{\alpha} \langle pq | \alpha \rangle + \sum_{\gamma} \varepsilon_{\gamma} \langle pqr \rangle \text{ を計算する。}$$

SCF=i (>1) のときはそれぞれ、

$$\sum_{\alpha} \Delta \rho_{\alpha}^{(i)} \langle pq | \alpha \rangle + \sum_{\gamma} \Delta \mu_{\gamma}^{(i)} \langle pqr \rangle$$

$$\sum_{\alpha} \Delta \rho_{\alpha}^{(i)} \langle pq | \alpha \rangle + \sum_{\gamma} \Delta \varepsilon_{\gamma}^{(i)} \langle pqr \rangle$$

を計算する。ここで $\Delta A^{(i)} = A^{(i)} - A^{(i-1)}$ であり (i) は SCF 回数を示す。

3 - 2 関連クラスの変更

現行のアルゴリズムにおいて DfEri は DfFockmatrix と DfTotalenergy において共通の

形で呼び出されている。よって、新クラス DfThreeindexintegrals から DfEri を呼び出す場合でも、DfEri クラスの変更はほとんど不要であった。

その一方で DfOverlap の場合は、Kohn-Sham 行列計算と全エネルギー計算のそれぞれに必要な展開係数が異なっていることより、各々の展開係数と、その答えを返す各々の行列を引数として一度に渡すように変更した。

3 - 3 使用メモリ

3 - 2 より、DfOverlap の計算をまとめて行う場合には、Kohn-Sham 行列計算に必要なハーフ行列と、全エネルギー計算に必要なハーフ行列が、各々 1 つずつ必要である。また本クラスでは DfEri と DfOverlap は別々に計算を行う。したがって新しいクラスでは合計 1 個の行列分のメモリが確保できればよい。

例えば 100 残基のタンパク質を計算する場合には 10000 × 10000 の行列となり 800MB のメモリが必要である。

上記の理論に基づいて ProteinDF の高速化を行った。実際にペプチド鎖の計算を行い、計算時間を比較した結果について、本発表で詳細に述べる。

4. 謝辞

本研究は文部科学省リサーチ・レポリエーション計画 (RR2002)、IT プログラム「戦略的基盤ソフトウェアの開発」の支援の下におこなわれました。

5. 参考文献

[1] F.Sato, Y.Shigemitsu, I.Okazaki, S.Yahiro, M.Fukue, S.Kozuru, H.Kashiwagi, Development of a new density functional program for all-electron calculation of proteins. Int. J. Quant. Chem. 63(1997)245-256.

[2] F.Sato, T.Yoshihiro, M.Era, H.Kashiwagi, Calculation of all-electron wavefunction of hemoprotein cytochrome c by density functional theory. Chem. Phys. Lett. 341(2001)645-651.